

RESEARCH

Open Access

# *In Silico* profiling of deleterious amino acid substitutions of potential pathological importance in haemophilia A and haemophilia B

George Priya Doss C<sup>†</sup>

## Abstract

**Background:** In this study, instead of current biochemical methods, the effects of deleterious amino acid substitutions in *F8* and *F9* gene upon protein structure and function were assayed by means of computational methods and information from the databases. Deleterious substitutions of *F8* and *F9* are responsible for Haemophilia A and Haemophilia B which is the most common genetic disease of coagulation disorders in blood. Yet, distinguishing deleterious variants of *F8* and *F9* from the massive amount of nonfunctional variants that occur within a single genome is a significant challenge.

**Methods:** We performed an *in silico* analysis of deleterious mutations and their protein structure changes in order to analyze the correlation between mutation and disease. Deleterious nsSNPs were categorized based on empirical based and support vector machine based methods to predict the impact on protein functions. Furthermore, we modeled mutant proteins and compared them with the native protein for analysis of protein structure stability.

**Results:** Out of 510 nsSNPs in *F8*, 378 nsSNPs (74%) were predicted to be 'intolerant' by SIFT, 371 nsSNPs (73%) were predicted to be 'damaging' by PolyPhen and 445 nsSNPs (87%) as 'less stable' by I-Mutant2.0. In *F9*, 129 nsSNPs (78%) were predicted to be intolerant by SIFT, 131 nsSNPs (79%) were predicted to be damaging by PolyPhen and 150 nsSNPs (90%) as less stable by I-Mutant2.0. Overall, we found that I-Mutant which emphasizes support vector machine based method outperformed SIFT and PolyPhen in prediction of deleterious nsSNPs in both *F8* and *F9*.

**Conclusions:** The models built in this work would be appropriate for predicting the deleterious amino acid substitutions and their functions in gene regulation which would be useful for further genotype-phenotype researches as well as the pharmacogenetics studies. These *in silico* tools, despite being helpful in providing information about the nature of mutations, may also function as a first-pass filter to determine the substitutions worth pursuing for further experimental research in other coagulation disorder causing genes.

**Keywords:** *In silico*, *F8*, *F9*, Haemophilia A, Haemophilia B

## Background

Hereditary haemophilias are the most frequently encountered recessive inherited disease of coagulation disorders in blood. Haemophilia A and Haemophilia B are X-linked inherited bleeding disorder caused by a decreased activity or lack of coagulation factor VIII

cofactor activity (haemophilia A) or coagulation factor IX enzyme activity (haemophilia B) due to heterogenous mutations in the *F8* and *F9* coding gene [1,2]. Factor VIII is a protein cofactor with no enzyme activity that, when activated, forms a complex with factor IXa serine protease on membrane surfaces. Upon activation, and in the presence of calcium ions and phospholipid surfaces, factor VIII and factor IX form an active complex, the tenase complex, which activates factor X during blood coagulation [3]. The *F8* gene maps to the distal end of the long arm of X-chromosome (Xq28) and spans 186

Correspondence: georgecp77@yahoo.co.in

<sup>†</sup> Contributed equally

Centre for Nanobiotechnology, Medical Biotechnology Division, School of Bio Sciences and Technology, VIT University, Vellore 632014, Tamil Nadu, India

kilo bases (kb) of genomic DNA. It consists of 26 exons and encodes a mature protein of 2,332 amino acids arranged within six domains organized as A1-A2-BA3-C1-C2 [4]. Its prevalence rate is estimated at 1:5,000-10,000 in men. Factor VIII circulates in the blood as a hetero dimer composed of two polypeptide chains: a light chain with a molecular weight of 80,000 Daltons (Da) and a heterogeneous heavy chain with a molecular weight varying between 90,000 and 200,000, Daltons (Da) both derived from the single peptide chain [5]. The *F9* gene is much smaller than *F8* maps to the distal end of the long arm of X-chromosome (Xq27) and spans 34 kb in length [6]. It contains 8 exons and encodes a glycoprotein of 415 amino acid residues, normally presents in plasma, which is an essential component of the clotting cascade [7]. It contains six major domains: signal peptide, propeptide, gla domain, two epidermal growth factor-like (EGF-like) domains, activation and catalytic domains [8]. The heterogeneous genetic diseases Haemophilia A & B, has been associated with missense mutations, nonsense mutations, gene deletions of varying size, insertions, inversions, and splice junction mutations and reported in Haemophilia A human database [9] and Haemophilia B human Database [2]. Classification of Haemophilia is based on plasma procoagulant levels, with persons with less than 1% active factor (< 0.01 IU/ml) are classified as having severe haemophilia, those with 1-5% active factor (0.01-0.05 IU/ml) have moderate haemophilia, and those with 5-40% of normal levels of active clotting factor (> 0.05- < 0.4 IU/ml) have mild haemophilia [10].

Recent advances in high-throughput genotyping and next generation sequencing have generated a tremendous amount of human genetic variation data, determining the effects of amino acid substitution will be the next challenge in mutation research. In the human genome single base substitutions called 'Single Nucleotide Polymorphisms' (SNPs) is the most frequent type of genetic variation. When SNPs occur in coding regions and produce amino acids change in the corresponding proteins, we name it as nonsynonymous single nucleotide polymorphisms (nsSNPs) [11]. Half of all genetic changes related to human diseases are attributed to nsSNP variants [12]. Differentiating deleterious nsSNPs (significant phenotypic consequences) from tolerant nsSNPs (without phenotypic change) are of great importance in understanding the genetic basis of haemophilias. This can be achieved by two general strategies: (i) carrier detection by linkage analysis (ii) *in silico* approach. Discriminating the types of nsSNPs in Mendelian disease genes, coupled with issues of statistical power, provide a compelling rationale for the application of a sequence-based approach to association studies rather than complete reliance on a map of

anonymous haplotypes [13]. Because genome-wide scans are still financially challenging, it is advantageous to prioritize variants that may affect the structure or function of expressed proteins. NsSNPs can be analyzed according to the biochemical severity of the amino acid substitution and its context within the protein sequence. In this context, Grantham matrix [14] predicts the effect of amino acid substitutions based on chemical properties, including polarity and molecular volume. Recently more sophisticated *in silico* programs were developed and made available on the World Wide Web [15-23]. They take into account, to various degrees, factors such as the general rules of protein chemistry (e.g., change in charge or in hydrophobicity, or helix-breaking residue), the three dimensional structure of the protein, and homologies in amino acid sequences among various species or related proteins. These tools made use of a variety of features such as information derived from protein sequences or from both sequence and structural information. Deleterious nsSNPs analyses for the *F8* and *F9* genes have not been estimated computationally till now, although they have received great attention from experimental researchers. To answer this question, in the absence of other experimental investigations, we tested empirical rule based methods PolyPhen (Polymorphism Phenotyping) and Sorts Intolerant From Tolerant (SIFT) [11,20], machine-learning approach I-Mutant 2.0 [21], UTRScan [22] and PupaSuite [23] were used for prioritization of high-risk SNPs in coding and non-coding regions (5' and 3' un-translated regions (UTR) SNPs). Based on the scores of SIFT, PolyPhen and I-Mutant, we identified the deleterious nsSNPs that are likely to affect the protein structure. In order to understand the molecular mechanism of disease, it is important to determine the impact of these mutations on the structure. We have identified the potential mutations, proposed modeled structure for the mutant proteins, and compared them with the native protein. We also analyzed native and mutant modeled protein for stability analysis, solvent accessibility and secondary structure analysis.

## Methods

### Extraction of SNP information

The SNPs information (Protein accession number (NP), amino acid position, SNP ID, UniProtKB/Swiss-Prot source ID, and mRNA accession number (NM) of *F8* and *F9* was retrieved from the NCBI dbSNP [24] and SWISS-Prot databases [25] for our computational analysis. The information on the effect and the relationship between the nsSNPs and Haemophilia A and Haemophilia B disease was compiled from *in vivo* and *in vitro* experiments according to PubMed, OMIM [26],

HAMSTeRS Database [9] and Haemophilia B [2] database and UniProtKB/Swiss-Prot databases.

#### Assessment of nsSNP functionality

Empirical rules are derived based on sequence information, structural information or both. These methods predict deleterious nsSNPs based on physicochemical properties [14], protein structure [27-29], and cross species conservation [28-30]. The SIFT [20] and PolyPhen server [11] are the two representatives for this purpose.

#### SIFT

SIFT program <http://blocks.fhcrc.org/sift/SIFT.html> uses "sequence homology to predict whether an amino acid substitution will affect protein function and hence, potentially alter phenotype". SIFT scores are classified as intolerant (0.00-0.05), potentially intolerant (0.051-0.10), borderline (0.101-0.20), or tolerant (0.201-1.00) [31,32]. The higher a tolerance index, the less functional impact a particular amino acid substitution is likely to have, as a higher tolerance index indicates that the position is less conserved across species.

#### PolyPhen

PolyPhen a multiple sequence alignment server that aligns sequences using structural information. PolyPhen performs the prediction through sequence-based characterization of the substitution site, calculation of position-specific independent count (PSIC) profile scores for two amino acid variants, and calculation of structural parameters and contacts. The higher a PSIC score difference, the higher functional impact a particular amino acid substitution is likely to have. Predictions of how a particular nsSNP may affect protein structure by PolyPhen 2.0 are assigned as "probably damaging" a score ( $\geq 2.000$ ) made with high confidence that the nsSNP should affect protein structure and/or function; "possibly damaging," score (1.500-1.999) where it may affect protein function and/or structure; and "benign," score (0.000-0.999) as most likely having no phenotypic effect.

#### I-Mutant

I-Mutant 2.0 available at <http://gpcr.biocomp.unibo.it/cgi/predictors/I-Mutant2.0/I-Mutant2.0.cgi>. is a support vector machine (SVM)-based tool for the automatic prediction of protein stability changes and stabilization centers upon single point mutations. I-Mutant 2.0 predictions are performed starting either from the protein structure or, more importantly, from the protein sequence [21]. The output file shows the predicted free energy change value or sign (DDG) which is calculated from the unfolding Gibbs free energy value of the mutated protein minus the unfolding Gibbs free energy value of the native type (kcal/mol). If the DDG value is

positive then the mutated protein will have high stability and vice versa for less stability. NsSNPs of *F8* and *F9* genes with experimental evidence of altered activity or disease association were considered as deleterious. The functional impact of the nsSNPs in *F8* and *F9* genes can be validated from the phenotypic data obtained from both *in vivo* and *in vitro* studies. Prediction accuracy of these computational methods was analyzed based on the positive findings from these benchmarking experiments obtained from HAMSTeRS Database and Haemophilia B database and UniProtKB/Swiss-Prot. nsSNPs predicted as "deleterious" and experimentally associated was considered as correct, while the prediction was defined as an error if such a deleterious nsSNP was predicted as tolerant. Concordance analysis between the functional consequences of each nsSNP of *F8* and *F9* genes predicted by SIFT, PolyPhen and I-Mutant were assessed using Spearman's rank correlation coefficient  $\rho$ .

#### Predicting the molecular phenotypic effects of deleterious SNPs

The PupaSuite 3.1 [23] are now synchronized to provide annotations for both noncoding and coding SNPs, as well as annotations for the SwissProt set of human disease mutations. PupaSuite finds all the SNPs mapping in locations that might cause a loss of functionality in the genes. PupasView [33] retrieves SNPs that could affect conserved regions that the cellular machinery uses for the correct processing of genes (intron/exon boundaries or exonic splicing enhancers).

#### Characterization of SNPs in regulatory untranslated regions

5' and 3' untranslated regions (UTR) of eukaryotic mRNAs are involved in many posttranscriptional regulatory pathways that control mRNA localization, stability and translation efficiency [34,35]. We used the program UTRScan <http://itbtools.ba.itb.cnr.it/utrscan> for this analysis. UTRScan looks for UTR functional elements by searching through user-submitted query sequences for the patterns defined in the UTRsite collection. Briefly, two or three sequences of each UTR SNP that have a different nucleotide at an SNP position are analyzed by UTRScan, which looks for UTR functional elements by searching through user-submitted sequence data for the patterns defined in the UTRsite and UTR databases. If different sequences for each UTR SNP are found to have different functional patterns, this UTR SNP is predicted to have functional significance.

#### Modeling nsSNP Locations on FVII and FIX

Structural analysis was performed based on the crystal structure of the protein for evaluating the structural stability of native and mutant protein. We used the

SAAPdb [36] and dbSNP to identify the protein coded by *F8* gene with PDB ID 2R7E[37] and *F9* gene with PDB ID 1RFN[38]. We also confirmed the mutation positions and the mutation residues from this server. These mutation positions and residues were in complete agreement with the results obtained with SIFT and PolyPhen programs. Based on the position of amino acids in the corresponding chains of the crystallized structures, the mutation analysis was performed using SWISSPDB viewer [39], and energy minimization was carried out using the program package GROMACS 4.0.5 [40] with Force field GROMOS96 43a1 [41]. The native and mutant proteins were solvated in cubic 0.9 nm of simple point charge (SPC) water molecules [42]. A periodic boundary condition was applied that the number of particles, pressure, and temperature was kept constant in the system. The system was neutralized by adding Na<sup>+</sup> and Cl<sup>-</sup> ions around the molecules to obtain electrically neutral system. The native and mutant structures were first minimized with steepest descent by 2000 steps and conjugated gradient by 3000. Computing total energy gives information about the protein structure stability. The deviation between the two structures is evaluated by their RMSD (root mean square deviation) values which could affect stability and functional activity [43]. By visualizing the position of the mutated amino acid residues; it is possible to suggest a physiochemical rationale for the effect on protein activity. The quality of 3D structure was assessed two programs: Verify 3D [44,45] and Prosa-Web [46].

#### Analyzing the effects of mutations on protein stability

We obtained the solvent accessibility information using program GETAREA [47] available at <http://curie.utmb.edu/getarea.html>. For a successful analysis of the relation between amino acid sequence and protein structure, an unambiguous and physically meaningful definition of secondary structure is essential. We obtained the information about secondary structures of the proteins using the program DSSP [48]. The prediction of solvent accessibility and secondary structure has been studied as an intermediate level for predicting the tertiary structure of proteins.

## Results

#### Predictions of deleterious and damaging coding nsSNPs

SNPs information for *F8* and *F9* was retrieved from dbSNP and cross verified with Swiss-prot database. For our investigations, we selected SNPs in nsSNPs and UTR (5' and 3') regions. Among the 675 nsSNPs, 510 nsSNPs (177 RefSNPs and 333 Swiss-Prot SNPs) and 165 nsSNPs (67 RefSNPs and 98 Swiss-Prot SNPs) were in *F8* and *F9*, 17 and 9 SNPs in mRNA of *F8* and *F9* were included in our analysis. We applied three *in*

*silico* tools SIFT, PolyPhen and I-Mutant 2.0 to predict the putative effect of each nsSNP on protein function.

#### F8

The protein sequences of 510 nsSNPs were submitted separately to the SIFT program to inspect its tolerance index. We identified a total of 378 nsSNPs (74%) that were scored as intolerant by SIFT. Approximately 197 nsSNPs (39%) exhibited SIFT scores of 0.0; 181 nsSNPs (35%) showed scores between 0.01-0.05; 9% of the variants (45 nsSNPs) have scores between 0.051-0.10 and the remaining 17% of the nsSNPs were classified as 'Tolerant' by SIFT. PolyPhen identified a total of 371 nsSNPs (73%) that were scored as damaging. 183 nsSNPs (36%) exhibited PolyPhen score of > 2.00, 188 nsSNPs (37%) have scores between 1.99-1.50, and 70 nsSNPs (13.5%) have scores between 1.49-1.25. Consequently, 69 nsSNPs (13.5%) were characterized as benign. We analyzed the nsSNPs finally by using I-Mutant server. 445 nsSNPs (87%) were found to be less stable and exhibited a DDG value ranging from -0.02 to -5.23 respectively. Approximately 296 nsSNPs (58%) showed a DDG value of ≤ -1.0; 169 nsSNPs (33%) showed a DDG value -1.01 to -2.00 and the remaining 45 nsSNPs (9%) showed a DDG value ≥ -2.01 respectively. Additional file 1: Table S1 represents the distribution of nsSNPs by SIFT, PolyPhen, and I-Mutant scores.

#### F9

Similarly, a total of 165 nsSNPs in *F9* gene were submitted to SIFT, PolyPhen and I-Mutant 2.0. Additional file 1: Table S1 represents the distribution of nsSNPs by SIFT PolyPhen and I-Mutant 2.0 scores. By SIFT, 90 nsSNPs (55%) showed a tolerance index score of 0.00; 39 nsSNPs (24%) showed scores between 0.01-0.05; 7 nsSNPs (4%) showed scores between 0.051-0.10 and the remaining 29 nsSNPs (17%) were classified as 'Tolerant' by SIFT. When PolyPhen was applied for prediction, 105 nsSNPs (64%) were scored as > 2.00; 26 nsSNPs (16%) exhibited scores between 1.99-1.50, 8 nsSNPs (4.5%) have scores between 1.49-1.25. Consequently, 69 nsSNPs (15.5%) were characterized as benign. By I-Mutant, 82 nsSNPs (49.6%) showed a DDG value of ≤ -1.0; 55 nsSNPs (33%) showed a DDG value -1.01 to -2.00, 13 nsSNPs (7.9%) showed a DDG value ≥ -2.01, 15 nsSNPs (9.1%) showed a DDG value < 0.00 respectively.

#### Correlation of computational methods in prediction of nsSNPs in haemophliacs

We combined scores of different prediction programs SIFT, PolyPhen, and I-Mutant 2.0 and found that this could significantly increase prediction performance nsSNPs analysis (Table 1). There are evidences to state combinatorial approach using different computational



**Table 1 Concordance analysis between SIFT and PolyPhen in the prediction of functional variants in F8 and F9**

SIFT		No of Variants		PolyPhen		No of Variants	
Scores	Prediction	F8	F9	Scores	Prediction	F8	F9
0.00-0.05	Intolerant	378	129	≥ 2.000	Probably damaging	183	105
0.051-0.10	Potentially intolerant	45	7	1.500-1.999	Possibly damaging	188	26
0.101-0.20	Borderline	43	8	1.25-1.49	Potentially damaging	70	8
0.201-1.00	Tolerant	44	21	1.00-1.24	Borderline	49	3
				0.000-0.999	Benign	20	23
<b>Total</b>		<b>510</b>	<b>165</b>	<b>Total</b>		<b>510</b>	<b>165</b>

methods performed well in increasing the accuracy in prediction of functional and deleterious nsSNPs [49]. Since a lower SIFT or I-Mutant 2.0 score indicate that the nsSNPs of interest would be more deleterious, whereas a higher PolyPhen score indicate that the nsSNPs of interest would be more deleterious. Among 510 nsSNPs in F8, 378 nsSNPs (74%) were predicted to be intolerant by SIFT, 371 nsSNPs (73%) were predicted to be damaging by PolyPhen and 445 nsSNPs (87%) as less stable by I-Mutant2.0. In F9, 129 nsSNPs (78%) were predicted to be intolerant by SIFT, 131 nsSNPs (79%) were predicted to be damaging by PolyPhen and 150 nsSNPs (90%) as less stable by I-Mutant2.0. By our analysis we found that I-Mutant outperformed SIFT and PolyPhen in prediction of deleterious nsSNPs in both F8 and F9. Most of these differences are likely the result of each method requiring a sufficient number and diversity of aligned sequences in order to make a prediction, each method using a different set of sequences and alignments. Our earlier analysis also shown individual tools correlate modestly with observed results, and that combining information from different tools may perform better in increasing the predictive accuracy in determining the functional impact of a given nsSNP [49]. In combination the nsSNPs which were predicted to be deleterious in causing an effect in the structure and function of the protein by SIFT, PolyPhen, and I-Mutant 2.0 correlated well experimental studies as shown in Table S1 [50-130].

#### Predictions of potential phenotypic effect in SNPs

Among 29 SNPs predicted by PupaSuite, 26 nsSNPs were found to disrupt Exon Splicing Enhancer and 3 nsSNPs were predicted to disrupt Exon Splicing Silencer as depicted in Additional file 1: Table S1 respectively. Four SNPs namely rs1803603, rs34683807, rs1396947, and rs5986887 were predicted to disrupt Exon Splicing Enhancer and SNP namely rs34700571 was predicted to disrupt Exon Splicing Silencer in untranslated region of F8 gene.

#### Functional SNPs in non-coding SNPs

Polymorphism in the 3'UTR region affects the gene expression by affecting the ribosomal translation of

mRNA or by influencing the RNA half-life. UTRResource was applied to prioritize 17 SNPs (F8) and 9 SNPs in (F9) UTR region. After comparing the functional elements for each UTR SNP, we found that only 5 SNPs were predicted to have functional significance. Four SNPs namely rs36101366, rs34683807, rs5986887, rs1396947 were related to functional pattern change of Upstream Open Reading Frame (UOF) and rs4487960 related to Polyadenylation Signal Upstream Open Reading Frame (uORF) in F8, and two SNPs with ID rs191483077 and rs186616567 were related to functional pattern change of Internal Ribosome Entry Site (IRES) in F9 respectively.

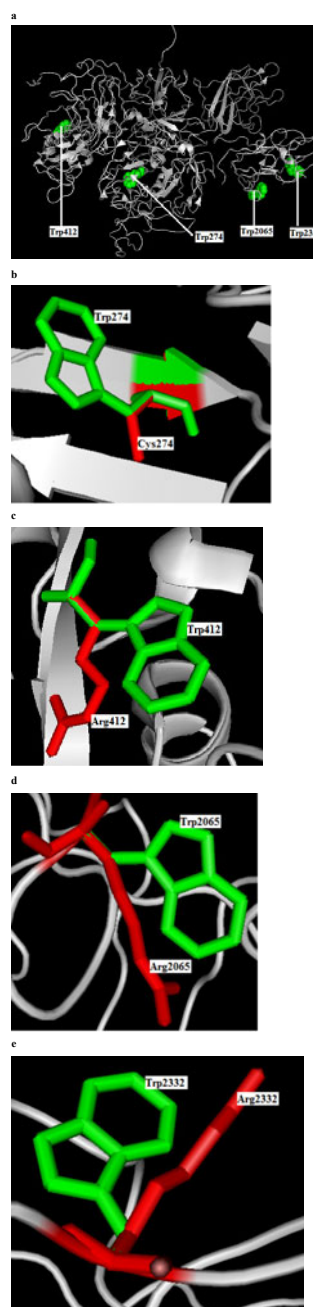
#### Structural analysis

Single amino acid mutations can significantly alter the stability of a protein structure. So, the knowledge of a protein's three-dimensional (3D) structure is essential for a full understanding of its functionality. Mapping the deleterious nsSNPs into protein structure information was obtained from dbSNP and SAAPdb. Available X-ray crystallized structures for the FVIII and FIX protein in Protein Data Bank with PDB ID code 2R7E (3.70 Å), and 2WPH (1.5 Å). Mutation analysis was performed based on the results obtained from highest SIFT, and PolyPhen scores. It is noted that rs34371500 (W274C), VAR\_028524 (W412R) in 'A' chain and rs28937299/rs137852455 (W2065R), and VAR\_028712 (W2332R) in 'B' chain of PDB ID 2R7E, showed the highest deleterious score of 0.00 (SIFT) and damaging scores (PolyPhen) ranging from 3.318 to 3.543 respectively in FVIII. Similarly in FIX, VAR\_006611 (W431R), and rs137852269 (W453R) showed the highest deleterious score of 0.00 (SIFT) and damaging scores (PolyPhen) of 4.434 and 4.632 respectively. For W431R and W453R mutation analysis was performed in the 'S' chain of the PDB ID 2WPH. The mutations for FVIII and FIX at their corresponding positions were performed by SWISS-PDB viewer independently to achieve modeled structures. Then, energy minimizations were performed by GROMACS 4.0.5 for the native type protein and the mutant type structures. Total energy and the RMSD

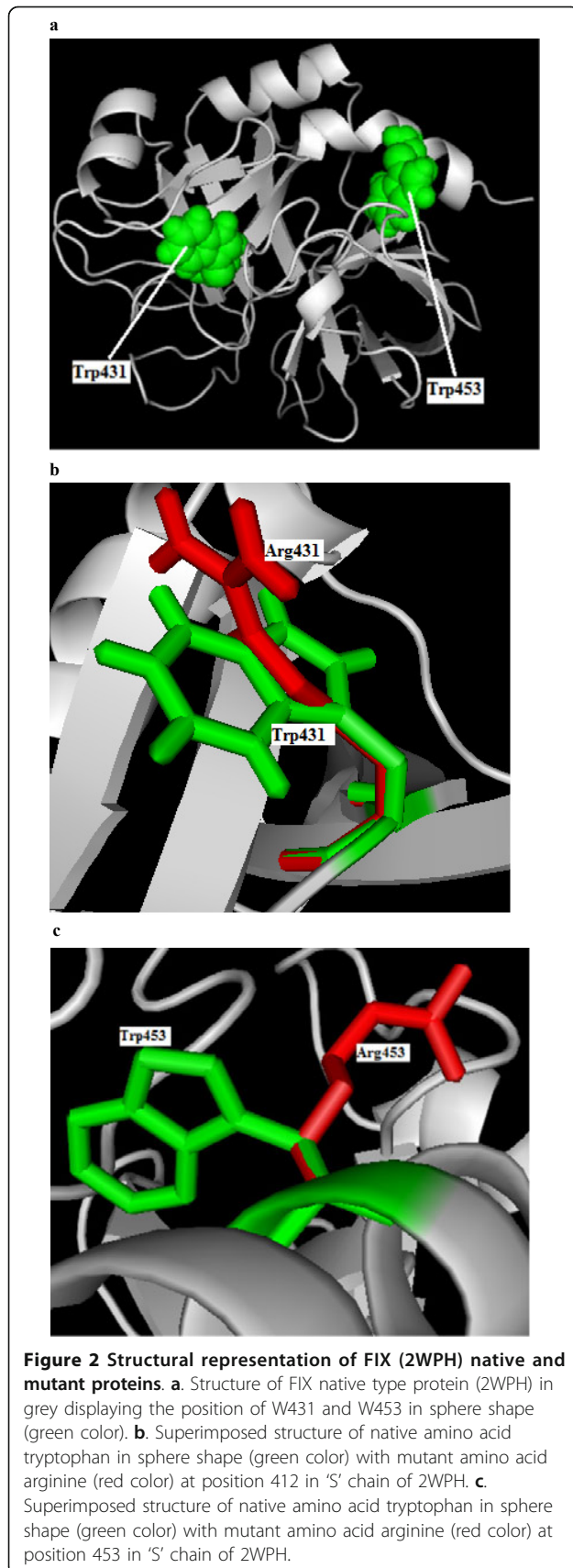
values between the native (2R7E and 2WPH) and the mutant amino acids were calculated. Higher the RMSD value more will be the deviation between native and mutant type structures and which in turn changes their functional activity. In this analysis found that the total energy for the mutant proteins W274C, W412R, W2065R, and W2332R following energy minimization was -97899.13, -98142.42, -98013.21 and -97013.21 kJ/mol when compared to native protein (2R7E) energy -98911.33 kJ/mol. The RMSD values were calculated between the native and mutant amino acids and showed 2.74 Å in W274C, 2.78 Å in W412R, 2.85 Å in W2065R and 2.91 Å in W2332R. The superimposed structures of the native protein with the four mutant type proteins are shown in Figure 1a-d respectively. These figures were drawn using PyMOL54 release 0.99 [131]. Similarly, we checked the total energy for mutant type structure W431R and W453R were found to be -81428.83 and -81694.21 when compared to native energy of -84591.35 kJ/mol. The RMSD values were calculated between the native and mutant amino acids and showed 2.94 Å in W431R, and 3.18 Å in W453R. The superimposed structures of the native protein with the four mutant type proteins are shown in Figure 2a-c, respectively. In W412R, W2065R, W2332R W431R and W453R there is change in drift of charge from non-polar to polar residue Substitution of positively charged arginine in place of neutral tryptophan may lead to disturbance in the interactions with other molecules or other parts of the proteins. These types of substitutions could introduce repulsive interactions between neighboring residues. Similarly we observed the potential effects of substitutions such as disruption of ligand binding site, disruption of annotated functional site, overpacking at buried site, contact with functional site, and hydrophobicity change at buried site in PolyPhen predictions. Solvent accessibilities and secondary structures of amino acid residues in the native and mutant proteins were analyzed by GETAREA and DSSP as shown in Table 2.

## Discussion

Predicting phenotypic consequences of nsSNPs by application of bioinformatics analysis may provide a good way to explore the function of nsSNPs and the relationship between nsSNPs and susceptibility to disease. The mutations causing haemophilia A and B have been localized and well characterized by several experimental studies. Most of the mutations in haemophilias leads to insufficient activity of the tenase complex, brought about either by a deficiency of coagulation factor VIII cofactor activity (haemophilia A) or coagulation factor IX enzyme activity (haemophilia B). Thus, it is not surprising that the two disorders are clinically



**Figure 1 Structural representation of FVIII (2R7E) native and mutant proteins. a.** Structure of FVIII native type protein (2R7E) in grey displaying the position of W274, W412, W2065 and W2232 in sphere shape (green color). **b.** Superimposed structure of native amino acid tryptophan in sphere shape (green color) with mutant amino acid cysteine (red color) at position 274 in 'A' chain of 2R7E. **c.** Superimposed structure of native amino acid tryptophan in sphere shape (green color) with mutant amino acid arginine (red color) at position 412 in 'A' chain of 2R7E. **d.** Superimposed structure of native amino acid tryptophan in sphere shape (green color) with mutant amino acid arginine (red color) at position 2065 in 'B' chain of 2R7E. **e.** Superimposed structure of native amino acid tryptophan in sphere shape (green color) with mutant amino acid arginine (red color) at position 22232 in 'B' chain of 2R7E.



similar because they both arise from perturbation of the same essential step in the process of fibrin generation. It is also clearly evident the molecular basis of the haemophilias is extremely diverse from the enormous number of mutations that have been elucidated so far. For this purpose, vast number of bioinformatic tools, based on recent findings from evolutionary biology (amino acid sequence), protein structure analysis, and computational biology may provide useful information in assessing the functional significance of SNPs have been proposed. In this study, we explored the relationship between prediction consequences of nsSNPs by computational approaches based on recent findings from evolutionary biology, protein structure research, and real phenotypes confirmed by experiments. The recent progress made in experimental 3D structure determination of FVIII and FIX by X-ray crystallography [37,38] and modeling studies [132] have made it possible to predict the effects of nsSNPs at structural level by mapping them on corresponding structures. The functional consequences of most SNPs *F8* and *F9* gene are still unknown, although some nsSNPs have been associated with X-linked inherited bleeding disorder. *In vivo* and *in vitro* studies on the function of nsSNPs have found that genetic mutations in *F8* and *F9* gene are responsible for Haemophilia A and Haemophilia B. There have been a quite lot of studies existing to validate the importance of single amino acid substitutions in Haemophilia A and Haemophilia B at activation cleavage sites [65,133,134], affecting factor VIII binding to von Willebrand factor [65,101,135] factor VIII secretion [92] and factor IX binding to factor VIII [136-138]. Recently Markoff by his homology modeling approach analyzed the impact of substitutions in loss of S-S bridges, thrombin activation site, gain/loss of H-bonds, cross-chain H-bonds, ionic bond and possible contact residue in FVIII [132]. The information regarding the involvement of mutations in Gla (γ-carboxyglutamic acid) domain, EGF1 for the N-terminal domain that binds calcium, EGF2 where it does not bind calcium, and a catalytic serine protease (SP) domain has been deposited in CoagMDB database [138]. The most commonly observed amino acid substitutions are Arg, Tyr, Phe and Cys in FVIII and FIX which play important role in altering the protein function. These amino acids are involved in protein folding dependent on disulfide bonds (Cys) and protein active or binding sites (Arg) [139,140]. These amino acids predicted deleterious by SIFT, PolyPhen and I-Mutant were in concordance with the experimental studies (Additional file 1: Table S1). It is becoming clear that implementation of the molecular evolutionary approach may be a powerful tool for prioritizing SNPs to be genotyped in future molecular epidemiological studies.



**Table 2 Solvent accessibilities and Secondary structure analysis in the native and mutant proteins**

Mutation Position	Solvent accessibility in the native and mutant proteins by GETAREA		Secondary structure analysis by DSSP
	Changed from exposed to buried	Changed from buried to exposed	
W274C	Thr (2), Tyr(5) Leu(7), Val(63), His(76), Ala(81), Pro(93), Ser(116), Ala(119), Glu(129), Lys(146), Tyr(155), Lys(185), Leu(187), Ala(194), Lys(213), Leu (217), Ile(310), Leu(319), Leu(327), Gln(335), Glu(340), Lys(344), Pro(349), Lys(399), Thr(400), Leu(419), Tyr(426), Lys(431), Lys(441), Ala(449), Thr(454), Lys(485), Arg(509), Gly(539), Thr(549), Ser(553), Glu(559), Glu(576), Asp(579), Gln(585), Lys(589), Val(592), Phe(598), Arg(602), Leu(614), Asn(631), Tyr(639), Ser(647), Ala(654), His(679), Leu(689), Val(697), Asn(713), Arg(719), Ser(728), Lys(732)	Trp(14), Leu(69), Gln(96), Val(99), Ser(138), Ser (176), Ser(179), Ile(405), Asn(299), Cys(329), Val (392), Arg(424), Tyr(473), Arg(503), Arg(546), Val(556), Pro(569), Gly(565), Asn(609), Gly(619), Ile(636), Trp(707)	T → H: Pro(86), Gly(89), Leu(203), Thr (208), Ser(428), Lys(518), Gly(638), His (1735), Thr(1763), Lys(1932), Asn(1934), Met(1945), Glu(2200) H → T:Lys(161), Lys(185), Leu(187), Leu (562), Asn(601), Glu(608), Arg(612), Ser (630), Asp(1865), Gly(2022) T → S:Trp(87), Leu(517), Asn(637), Asn (1772)
W2065R	Asp(1260), Gln(1336), Leu(1481), Ala(1610), Ile (1698), Tyr(1699), Arg(1708), Thr(1714), Glu (1723), Arg(1740), Gly(1769), Leu(1775), Ile (1782), Arg(1800), His(1867), Leu(1882), Trp (1908), Ala(1939), Asn(1941), Met(1945), Arg (1960), Ser(1968), Asn(1971), Phe(1982), Met (2007), Arg(2016), Ser(2082), Arg(2169)	Asn(1460), Gln(1705), Gly(1779), Leu(1808), Val (1876), Glu(1884), Glu(1904), Met(1842), Ile (1901), Tyr(1909), Thr(2015),	T → H:), His(1735), Thr(1763), Lys(1932), Asn(1934), Met(1945), Glu(2200) H → T: Asp(1865), His(1867), Asp(2206) T → S: Asn(1772)
W2248C	Asp(1260), Gln(1336), Asn(1460), Leu(1481), Ala (1610), Ile(1698), Tyr(1699), Arg(1708), Thr (1714), Glu(1723), Arg(1740), Gly(1769), Leu (1775), Ile(1782), Arg(1800), His(1867), Val(1876), Leu(1882), Trp(1908), Ala(1939), Asn(1941), Met (1945), Arg(1960), Ser(1968), Asn(1971), Phe (1982), Met(2007), Arg(2016), Gly(2028), Ala (2070)Ser(2082), Arg(2169)	Asn(1460), Gln(1705), Gly(1779), Leu(1808), Val (1876), Glu(1884), Glu(1904), Met(1842), Ile (1901), Tyr(1909), Thr(2015), Gly(2022), Ala (2070)	T → H:), His(1735), Thr(1763), Lys(1932), Asn(1934), Met(1945), Glu(2200) H → T: Asp(1865), Gly(2022)

T-Turn, H-Helix, S-Strand

Moreover, from an evolutionary perspective, SNPs altering a conserved amino acid site are more likely to have functional importance. Computational tools like SIFT and PolyPhen are able to predict 90% of damaging SNPs. Several groups also validated these algorithms has, however, come from benchmarking studies based on the analysis of “known” deleterious substitutions annotated in databases, such as SwissProt. In such studies, PolyPhen and SIFT has been shown to successfully predict the effect of over 80% of amino acid substitutions [31,32,141]. In this study, we first surveyed previous publications and submitted mutations in database associated with Haemophilia A and Haemophilia B, the most extensively examined bleeding disorder. In this present study three different widely-used computational tools were employed for determining the functional significance of nsSNPs. First, we included functional scores from SIFT, PolyPhen and I-Mutant tools, each of which employs fundamentally different algorithms that can be used to determine the functionality of the same nsSNPs. Proteins with mutations do not always have 3D structures that are solved and deposited in PDB. Therefore, it is necessary to construct 3D models using homology modeling by locating the variation in 3D. This is a simple way of

detecting what kind of adverse effects that a mutation can have on a protein.

Based on the SIFT, PolyPhen, and I-Mutant scores and availability of 3D structures, structure analysis was carried out with the major mutation that occurred in the native protein coded by *F8* and *F9*. The total energy and RMSD value of mutant structures W274C, W412R, W2065R, W2332R, W431R, and W453R were calculated. Correlations between SIFT (Sensitivity 83%), PolyPhen (Sensitivity 82%) and I Mutant (Sensitivity 80%) were calculated from raw scores rather than the arbitrarily defined categories. There was a significant correlation between the predictions obtained using SIFT and PolyPhen algorithms in both *F8* ( $\rho = -0.59$ ) and *F9* ( $\rho = -0.576$ ); while the correlation between the predictions obtained using PolyPhen and I Mutant in *F8* ( $\rho = -0.86$ ) and *F9* ( $\rho = -0.63$ ) were much higher. A positive correlation was observed with SIFT and I- Mutant score for *F8* ( $\rho = 0.30$ ) and *F9* ( $\rho = 0.31$ ). We have shown that our data suggests that different tools correlate modestly with observed results, and that combining information from a variety of tools may significantly increase the predictive power for determining the functional impact of a given nsSNP.



## Conclusion

In conclusion, from our *in silico* analysis it is very difficult to determine whether the notable differences exist in the performance of these methods in predicting deleterious nsSNPs in *F8* and *F9*. The variation in the prediction might be due to the difference in features utilized by the methods or the training datasets. There is no single literature stating a single *in silico* method can aid in better prediction. Our *in silico* analysis coincides with previous analysis performed by other groups stating that combining information obtained from various methods can increase prediction performance. The overall strategy of our study was to prioritize the functional nsSNPs, map as many structural mutations as possible, find general patterns to analyze 3D mutations with respect to protein function and evaluate regulatory variants using many *in silico* analysis methods as possible. Based on these analyses, we try to determine the relationship between the disease-related mutations and structural properties of proteins in haemophiliacs.

## Additional material

**Additional file 1: Table S1.** List of nsSNPs found to be functionally significant by SIFT, PolyPhen, I-Mutant and PupaSuite.

## Abbreviations

ESE: Exon splicing enhancer; ESS: Exon splicing silencer; NCBI: National center for biotechnology information; nsSNPs: Non-synonymous single nucleotide polymorphisms; OMIM: Online mendelian inheritance in man; PSIC: Position specific independent count; RMSD: Root mean square deviation; SIFT: Sorting intolerance from tolerance; SNP: Single nucleotide polymorphisms; UTR: Untranslated region.

## Acknowledgements

The authors thank the management of VIT University for providing the facilities to carry out this work.

## Authors' contributions

CGPD collected the SNP data from the databases, analyzed the SNPs using different algorithms, predicted and the deleterious SNPs. CGPD also carried out the modeling analysis and drafted the manuscript.

## Competing interests

The author declares that they have no competing interests.

Received: 29 April 2011 Accepted: 16 March 2012

Published: 16 March 2012

## References

1. Renault NK, Dyack S, Dobson MJ, Costa T, Lam WL, Greer WL: **Heritable skewed X-chromosome inactivation leads to haemophilia A expression in heterozygous females.** *Eur J Hum Genet* 2007, **15(6)**:628-637.
2. Giannelli F, Green PM, Sommer SS, Poon MC, Ludwig M, Schwaab R, Reitsma PH, Goossens M, Yoshioka A, Figueiredo MS, Brownlee GG: **Haemophilia B: database of point mutations and short additions and deletions.** *Nucleic Acids Res*, 7 1998, **26(1)**:265-268.
3. Bowen DJ: **Haemophilia A and haemophilia B: molecular insights.** *Mol Pathol* 2002, **55**:127-440.
4. Vehar GA, Keyt B, Eaton D, Rodriguez H, O'Brien DP, Rotblat F, Oppermann H, Keck R, Wood WI, Harkins RN, Tuddenham EGD, Lawn RM, Dapton DJ: **Structure of human factor VIII.** *Nature* 1984, **312**:337-42.
5. Oldenburg J, Ananyeva NM, Saenko EL: **Molecular basis of haemophilia A.** *Haemophilia* 2004, **10(4)**:133-139.
6. Ivaskевичius V, Jurgutis R, Rost S, Müller A, Schmitt C, Wulff K, Herrmann FH, Müller CR, Schwaab R, Oldenburg J: **Lithuanian haemophilia A and B registry comprising phenotypic and genotypic data.** *Br J Haematol* 2001, **112**:1062-70.
7. Roberts HR: **Molecular biology of haemophilia B.** *Thromb Haemost* 1993, **70**:1-9.
8. Yoshitake S, Shach BG, Foster DC, Davie EW, Kurachi K: **Nucleotide sequence of the gene for human factor IX (antithemophilic factor B).** *Biochemistry* 1985, **24**:3736-3750.
9. **HAMSTeRS Database.** [http://hadb.org.uk/].
10. White GC, Rosendaal F, Aledort LM, Lusher JM, Rothschild C, Ingerslev J, On behalf of the Factor VIII and Factor IX Subcommittee: **Definitions in Haemophilia. Recommendation of the Scientific Subcommittee on Factor VIII and Factor IX of the Scientific and Standardization Committee of the International Society on Thrombosis and Haemostasis.** *Thromb Haem* 2001, **85**:560.
11. Ramensky V, Bork P, Sunyaev S: **Human non-synonymous SNPs: server and survey.** *Nucleic Acids Res* 2002, **30(17)**:3894-3900.
12. Stenson PD, Mort M, Ball EV, Howells K, Phillips AD, Thomas NS, Cooper DN: **The human gene mutation database: 2008 update.** *Genome Medicine* 2009, **1**:13.
13. Botstein D, Risch N: **Discovering genotypes underlying human phenotypes: past successes for Mendelian disease, future approaches for complex disease.** *Nat Genet* 2003, **33**:228-237.
14. Grantham R: **Amino acid difference formula to help explain protein evolution.** *Science* 1974, **185**:862-864.
15. Bromberg Y, Rost B: **SNAP: predict effect of non-synonymous polymorphisms on function.** *Nucleic Acids Res* 2007, **35**:3823-3835.
16. Ng PC, Henikoff S: **SIFT: predicting amino acid changes that affect protein function.** *Nucleic Acids Res* 2003, **31**:3812-3814.
17. Riva A, Kohane IS: **SNPper: retrieval and analysis of human SNPs.** *Bioinformatics* 2002, **18**:1681-1685.
18. Stitzel NO, Binkowski TA, Tseng YY, Kasif S, Liang J: **topoSNP: a topographic database of non-synonymous single nucleotide polymorphisms with and without known disease association.** *Nucleic Acids Res* 2004, **1(32)**:D520-D522.
19. Joshua SK, Yan Z, Colin W, Zemin Z: **CanPredict: a computational tool for predicting cancer-associated missense mutations.** *Nucleic Acids Res* 2007, **35**:W595-W598.
20. Kumar P, Henikoff S, Ng PC: **Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm.** *Nat Protoc* 2009, **4**:1073-1081.
21. Capriotti E, Fariselli P, Casadio R: **I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure.** *Nucleic Acids Res* 2005, **33**:W306.
22. Grillo G, Turi A, Licciulli F, Mignone F, Liuni S, Banfi S, Gennarino VA, Horner DS, Pavesi G, Picardi E, Pesole G: **UTRdb and UTRsite (RELEASE 2010): a collection of sequences and regulatory motifs of the untranslated regions of eukaryotic mRNAs.** *Nucleic Acids Res* 2010, **38**:D75-D80.
23. Reumers J, Conde L, Medina I, Maurer-Stroh S, Van Durme J, Dopazo J, Rousseau F, Schymkowitz J: **Joint annotation of coding and non-coding single nucleotide polymorphisms and mutations in the SNPeff and PupaSuite databases.** *Nucleic Acids Res* 2008, **36**:D825-D829.
24. Eric WS, Tanya B, Dennis AB, Evan B, Stephen HB, Kathi C, Vyacheslav C, Deanna MC, Michael D, Scott F, Michael F, Ian MF, Lewis YG, Wolfgang H, Yuri K, David L, David JL, Zhiyong L, Thomas LM, Tom M, Donna RM, Aron MB, Vadim M, Ilene M, James O, Anna P, Lon P, Kim DP, Gregory DS, Edwin S, Stephen TS, Martin S, Karl S, Douglas S, Alexandre S, Grigory S, Tatiana AT, Lukas W, Yanli W, W JW, Eugene Y, Jian Y: **Database resources of the National Center for Biotechnology Information.** *Nucleic Acids Res* 2011, **39**:D38-D51.
25. Amos B, Rolf A: **The SWISS-PROT Protein Sequence Data Bank and Its New Supplement TRMBL.** *Nucleic Acids Res* 1996, **24(1)**:21-25.

26. Amberger J, Bocchini CA, Scott AF, Hamosh A: **McKusick's Online Mendelian Inheritance in Man (OMIM)**. *Nucleic Acids Res* 2009, **37**: D793-D796.
27. Chasman D, Adams RM: **Predicting the functional consequences of nonsynonymous single nucleotide polymorphisms: structure-based assessment of amino acid variation**. *J Mol Biol* 2001, **307**:683-706.
28. Sunyaev S, Ramensky V, Koch I, Lathe W, Kondrashov A, Bork P: **Prediction of deleterious human alleles**. *Hum Mol Genet* 2001, **10**:591-597.
29. Wang Z, Moulton J: **SNPs, protein structure, and disease**. *Hum Mutat* 2001, **17**:263-270.
30. Burke DF, Worth CL, Priego E, Cheng T, Smink LJ, Todd JA, Blundell TL: **Genome bioinformatic analysis of nonsynonymous SNPs**. *BMC Bioinformatics* 2007, **8**:301-315.
31. Xi T, Jones IM, Mohrenweiser HW: **Many amino acid substitution variants identified in DNA repair genes during human population screenings are predicted to impact protein function**. *Genomics* 2005, **83**:970-979.
32. Ng PC, Henikoff S: **Predicting deleterious amino acid substitutions**. *Genome Res* 2001, **11**:863-874.
33. Conde L, Vaquerizas JM, Ferrer-Costa C, Orozco M, Dopazo J: **PupasView: a visual tool for selecting suitable SNPs, with putative pathological effect in genes, for genotyping Purposes**. *Nucleic Acids Res* 2005, **33**:W501-W505.
34. Sonenberg N: **mRNA translation: influence of the 5' and 3' untranslated Regions**. *Curr Opin Genet* 1994, **4**(2):310-15.
35. Nowak R: **Mining treasures from 'junk DNA'**. *Science* 1994, **263**:608-610.
36. Hurst JM, McMillan LE, Porter CT, Allen J, Fakorede A, Martin AC: **The SAAPdb web resource: a large-scale structural analysis of mutant proteins**. *Hum Mutat* 2009, **30**(4):616-624.
37. Shen BW, Spiegel PC, Chang CH, Huh JW, Lee JS, Kim J, Kim YH, Stoddard BL: **The tertiary structure and domain organization of coagulation factor VIII**. *Blood* 2008, **111**(3):1240-1247.
38. Hopfner KP, Lang A, Karcher A, Sichler K, Kopetzki E, Brandstetter H, Huber R, Bode W, Engh RA: **Coagulation factor IXa: the relaxed conformation of Tyr99 blocks substrate binding**. *Structure* 1999, **7**(8):989-996.
39. Kaplan W, Littlejohn TG: **Swiss-PDB Viewer (Deep View)**. *Brief Bioinform* 2001, **2**(2):195-197.
40. Hess B, Kutzner C, van der Spoel D, Lindahl E: **GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation**. *J Chem Theory Comput* 2008, **4**:435-447.
41. Van Gunsteren WF, Billeter SR, Eising AA, Hunenberger PH, Kruger P, Mark AE, Scott WRP, Tironi IG: **Biomolecular Simulation: The GROMOS96 Manual and User Guide**. vdf Hochschulverlag AG an der ETH Zurich and BIOMOS b.v. Zurich, Groningen; 1996.
42. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML: **Comparison of simple potential functions for simulating liquid water**. *J Chem Phys* 1983, **79**:926-935.
43. Varfolomeev SD, Uporov IV, Fedorov EV: **Bioinformatics and molecular modeling in chemical enzymology. Active sites of hydrolases**. *Biochemistry* 2002, **67**(10):1099-1108.
44. Bowie JU, Luthy R, Eisenberg D: **A method to identify protein sequences that fold into a known three-dimensional structure**. *Science* 1991, **253**:164-170.
45. Luthy R, Bowie JU, Eisenberg D: **Assessment of protein models with three dimensional profiles**. *Nature* 1992, **356**:83-85.
46. Wiederstein M, Sippl MJ: **ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins**. *Nucleic Acids Res* 2007, **35**:W407-W410.
47. Fraczkiewicz R, Braun W: **Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules**. *J Comp Chem* 1998, **19**:319-333.
48. Kabsch W, Sander C: **Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features**. *Biopolymers* 1998, **22**:2577-2637.
49. Rajith B, George Priya Doss C: **Path to facilitate the prediction of functional amino acid substitutions in red blood cell disorders - a computational approach**. *PLoS ONE* 2011, **6**(9):e24607.
50. Strmecki L, Benedik-Dolnicar M, Vouk K, Komel R: **Screen of 55 Slovenian haemophilia A patients: identification of 2 novel mutations (S-1R and IVS23+1G > A) and discussion of mutation spectrum**. *Mutation in brief no. 241*. Online. *Hum Mutat* 1999, **13**(5):413.
51. Cutler JA, Mitchell MJ, Smith MP, Savidge GF: **The identification and classification of 41 novel mutations in the factor VIII gene (F8C)**. *Hum Mutat* 2002, **19**:274-278.
52. Waseem NH, Bagnall R, Green PM, Giannelli F: **Start of UK confidential haemophilia A database: analysis of 142 patients by solid phase fluorescent chemical cleavage of mismatch**. Haemophilia Centres. *Thromb Haemost* 1999, **81**(6):900-905.
53. Becker J, Schwaab R, Moeller-Taube A, Schwaab U, Schmidt W, Brackmann HH, Grimm T, Olek K, Oldenburg J: **Characterization of the factor VIII defect in 147 patients with sporadic Haemophilia A: family studies indicate a mutation type-dependent sex ratio of mutation frequencies**. *Am J Hum Genet* 1996, **58**:657-670.
54. Diamond C, Kogan S, Levinson B, Gitschier J: **Amino acid substitutions in conserved domains of factor VIII and related proteins: study of patients with mild and moderately severe haemophilia A**. *Hum Mutat* 1992, **1**:248-257.
55. Tavassoli K, Eigel A, Pollmann H, Horst J: **Mutational analysis of ectopic factor VIII transcripts from haemophilia A patients: identification of cryptic splice site, exon skipping and novel point mutations**. *Hum Genet* 1997, **100**(5-6):508-511.
56. Valleix S, Vinciguerra C, Lavergne J-M, Leuer M, Delpech M, Negrier C: **Skewed X-chromosome inactivation in monozygotic diamniotic twin sisters results in severe and mild haemophilia A**. *Blood* 2002, **100**:3034-3036.
57. Liu ML, Nakaya S, Thompson AR: **Non-inversion factor VIII mutations in 80 haemophilia A families including 24 with alloimmune responses**. *Thromb Haemost* 2002, **87**(2):273-276.
58. Timur AA, Guergey A, Aktuglu G, Kavakli K, Canatan D, Olek K, Caglayan SH: **Molecular pathology of haemophilia A in Turkish patients: identification of 36 independent mutations**. *Haemophilia* 2001, **7**:475-481.
59. Goodeve AC, Williams I, Bray GL, Peake IR: **Relationship between factor VIII mutation type and inhibitor development in a cohort of previously untreated patients treated with recombinant factor VIII (Recombinate)**. *Recombinant PUP Study Group*. *Thromb Haemost* 2000, **83**(6):844-848.
60. Bicocchi MP, Pasino M, Lanza T, Bottini F, Boeri E, Mori PG, Molinari AC, Rosano C, Acquila M: **Analysis of 18 novel mutations in the factor VIII gene**. *Br J Haematol* 2003, **122**:810-817.
61. Bidichandani SI, Lanyon WG, Shiach CR, Lowe GD, Connor JM: **Detection of mutations in ectopic factor VIII transcripts from nine haemophilia A patients and the correlation with phenotype**. *Hum Genet* 1995, **95**(5):531-538.
62. Vencesla A, Corral-Rodriguez MA, Baena M, Cornet M, Domenech M, Baiget M, Fuentes-Prior P, Tizzano EF: **Identification of 31 novel mutations in the F8 gene in Spanish haemophilia A patients: structural analysis of 20 missense mutations suggests new intermolecular binding sites**. *Blood* 2008, **111**:3468-3478.
63. Bauduer F, Ducout L, Bendriss P, Falaises B, Lavergne JM: **Mild haemophilia A discovered in a previously multi-operated 73-year-old man: characterization of a new mutation**. *Haemophilia* 2001, **7**(4):419-421.
64. Liu M, Murphy MEP, Thompson AR: **A domain mutations in 65 haemophilia A families and molecular modelling of dysfunctional factor VIII proteins**. *Br J Haematol* 1998, **103**:1051-1060.
65. Higuchi M, Kazazian HH Jr, Kasch L, Warren TC, McGinniss MJ, Phillips JA, Kasper C, Janco R, Antonarakis SE: **Molecular characterization of severe Haemophilia A suggests that about half the mutations are not within the coding regions and splice junctions of the factor VIII gene**. *Proc Natl Acad Sci* 1991, **88**:7405-7409.
66. Bicocchi MP, Pasino M, Lanza T, Bottini F, Molinari AC, Caprino D, Rosano C, Acquila M: **Small FVIII gene rearrangements in 18 haemophilia A patients: five novel mutations**. *Am J Hematol* 2005, **78**(2):117-122.
67. Arruda VR, Pieneman WC, Reitsma PH, Deutz-Terlouw PP, Annichino-Bizzacchi JM, Briet E, Costa FF: **Eleven novel mutations in the factor VIII gene from Brazilian haemophilia A patients**. *Blood* 1995, **86**:3015-3020.
68. Maugard C, Tuffery S, Aguilar-Martinez P, Schved JF, Gris JC, Demaille J, Claustres M: **Protein truncation test: detection of severe haemophilia a mutation and analysis of factor VIII transcripts**. *Hum Mutat* 1998, **11**(1):18-22.
69. Theophilus BDM, Enayat MS, Williams MD, Hill FGH: **Site and type of mutations in the factor VIII gene in patients and carriers of haemophilia A**. *Haemophilia* 2001, **7**:381-391.

70. Leuer M, Oldenburg J, Lavergne JM, Ludwig M, Fregin A, Eigel A, Ljung R, Goodeve A, Peake I, Olek K: **Somatic mosaicism in Haemophilia A: a fairly common event.** *Am J Hum Genet* 2001, **69**:75-87.
71. Klopp N, Oldenburg J, Uen C, Schneppenheim R, Graw J: **11 haemophilia A patients without mutations in the factor VIII encoding gene.** *Thromb Haemost* 2002, **88**(2):357-360.
72. Chan V, Pang A, Chan TPT, Chan VW-Y, Chan TK: **Molecular characterization of haemophilia A in southern Chinese.** *Br J Haematol* 1996, **93**:451-456.
73. Albanez S, Ruiz-Saez A, Boadas A, De Bosch N, Porco A: **Identification of factor VIII gene mutations in patients with severe haemophilia A in Venezuela: identification of seven novel mutations.** *Haemophilia* 2011, **17**:913-918.
74. Vidal F, Farssac E, Altisent C, Puig L, Gallardo D: **Rapid haemophilia A molecular diagnosis by a simple DNA sequencing procedure: identification of 14 novel mutations.** *Thromb Haemost* 2001, **85**(4):580-583.
75. Mazurier C, Parquet-Gernez A, Gaucher C, Lavergne J-M, Goudemand J: **Factor VIII deficiency not induced by FVIII gene mutation in a female first cousin of two brothers with haemophilia A.** *Br J Haematol* 2002, **119**:390-392.
76. Möller-Morlang K, Tavassoli K, Eigel A, Pollmann H, Horst J: **Mutational-screening in the factor VIII gene resulting in the identification of three novel mutations, one of which is a donor splice mutation. Mutations in brief no. 245.** Online. *Hum Mutat* 1999, **13**(6):504.
77. Pieneman WC, Deutz-Terlouw PP, Reitsma PH, Briet E: **Screening for mutations in haemophilia A patients by multiplex PCR-SSCP, Southern blotting and RNA analysis: the detection of a genetic abnormality in the factor VIII gene in 30 out of 35 patients.** *Br J Haematol* 1995, **90**:442-449.
78. Williams IJ, Abuzenadah A, Winship PR, Preston FE, Dolan G, Wright J, Peake IR, Goodeve AC: **Precise carrier diagnosis in families with haemophilia A: use of conformation sensitive gel electrophoresis for mutation screening and polymorphism analysis.** *Thromb Haemost* 1998, **79**(4):723-726.
79. Yenchitsomanus P, Akkarapatumwong V, Pung-Amritt P, Intorasoot S, Thanootarakul P, Oranwiroon S, Veerakul G, Mahasandana C: **Genotype and phenotype of haemophilia A in Thai patients.** *Haemophilia* 2003, **9**:179-186.
80. Freson K, Peerlinck K, Aguirre T, Arnout J, Vermeylen J, Cassiman JJ, Matthijs G: **Fluorescent chemical cleavage of mismatches for efficient screening of the factor VIII gene.** *Hum Mutat* 1998, **11**(6):470-479.
81. Morichika S, Shima M, Kamisue S, Tanaka I, Imanaka Y, Suzuki H, Shibata H, Pemberton S, Gale K, McVey J, Tuddenham EGD, Yoshioka A: **Factor VIII gene analysis in Japanese CRM-positive and CRM-reduced haemophilia A patients by single-strand conformation polymorphism.** *Br J Haematol* 1997, **98**:901-906.
82. Youssoufian H, Wong C, Aronis S, Platokoukis H, Kazazian HH Jr, Antonarakis SE: **Moderately severe haemophilia A resulting from Glu- > Gly substitution in exon 7 of the factor VIII gene.** *Am J Hum Genet* 1988, **42**:867-871.
83. Tagariello G, Belvini D, Salvato R, Are A, De Biasi E, Goodeve A, Davoli P: **Experience of a single Italian center in genetic counseling for haemophilia: from linkage analysis to molecular diagnosis.** *Haematologica* 2000, **85**:525-529.
84. Mazurier C, Gaucher C, Jorieux S, Parquet-Gernez A: **Mutations in the FVIII gene in seven families with mild haemophilia A.** *Br J Haematol* 1997, **96**(2):426-427.
85. Habart D, Kalabova D, Novotny M, Vorlova Z: **Thirty-four novel mutations detected in factor VIII gene by multiplex CSGE: modeling of 13 novel amino acid substitutions.** *J Thromb Haemost* 2003, **1**:773-781.
86. Bogdanova N, Lemcke B, Markoff A, Pollmann H, Dworniczak B, Eigel A, Horst J: **Seven novel and four recurrent point mutations in the factor VIII (F8C) gene.** *Hum Mutat* 2001, **18**:546-546.
87. Citron M, Godmilow L, Ganguly T, Ganguly A: **High throughput mutation screening of the factor VIII gene (F8C) in haemophilia A: 37 novel mutations and genotype-phenotype correlation.** *Hum Mutat* 2002, **20**:267-274.
88. Kogan S, Gitschier J: **Mutations and a polymorphism in the factor VIII gene discovered by denaturing gradient gel electrophoresis.** *Pro Natl Acad Sci* 1990, **87**:2092-2096.
89. Frusconi S, Passerini I, Girolami F, Masieri M, Linari S, Longo G, Morfini M, Torricelli F: **Identification of seven novel mutations of F8C by DHPLC.** *Hum Mutat* 2002, **20**(3):231-232.
90. Tavassoli K, Eigel A, Dworniczak B, Valtseva E, Horst J: **Identification of four novel mutations in the factor VIII gene: three missense mutations (E1875G, G2088S, I2185T) and a 2-bp deletion (1780delTC).** *Hum Mutat* 1998, **1**:S260-S262.
91. Hill M, Deam S, Gordon B, Dolan G: **Mutation analysis in 51 patients with haemophilia A: report of 10 novel mutations and correlations between genotype and clinical phenotype.** *Haemophilia* 2005, **11**:133-141.
92. Roelse JC, De Laaf RT, Timmermans SM, Peters M, Van Mourik JA, Voorberg J: **Intracellular accumulation of factor VIII induced by missense mutations Arg593->Cys and Asn618->Ser explains cross-reacting material-reduced haemophilia A.** *Br J Haematol* 2000, **108**(2):241-246.
93. Hay CRM, Ludlam CA, Colvin BT, Hill FGH, Preston FE, Wasseem N, Bagnall R, Peake IR, Berntorp E, Mauser Bunschoten EP, Fijnvandraat K, Kasper CK, White G, Santagostino E: **Factor VIII inhibitors in mild and moderate-severity haemophilia A.** *Thromb Haemost* 1998, **79**:762-766.
94. Rudzki Z, Duncan EM, Casey GJ, Neumann M, Favaloro EJ, Lloyd JV: **Mutations in a subgroup of patients with mild haemophilia A and a familial discrepancy between the one-stage and two-stage factor VIII:C methods.** *Br J Haematol* 1996, **94**:400-406.
95. McGinniss MJ, Kazazian HH Jr, Hoyer LW, Bi L, Inaba H, Antonarakis SE: **Spectrum of mutations in CRM-positive and CRM-reduced haemophilia A.** *Genomics* 1993, **15**(2):392-8.
96. Paynton C, Sarkar G, Sommer SS: **Identification of mutations in two families with sporadic haemophilia A.** *Hum Genet* 1991, **87**:397-400.
97. Economou EP, Kazazian HH Jr, Antonarakis SE: **Detection of mutations in the factor VIII gene using single-stranded conformational polymorphism (SSCP).** *Genomics* 1992, **13**:909-911.
98. Nafa K, Baudis M, Deburgrave N, Bardin JM, Sultan Y, Kaplan JC, Delpech M: **A novel mutation (Arg->Leu in exon 18) in factor VIII gene responsible for moderate Haemophilia A.** *Hum Mutat* 1992, **1**:77-78.
99. Cai XH, Wang XF, Dai J, Fang Y, Ding QL, Xie F, Wang HL: **Female haemophilia A heterozygous for a de novo frameshift and a novel missense mutation of factor VIII.** *J Thromb Haemost* 2006, **4**(9):1969-1974.
100. Liu ML, Shen BW, Nakaya S, Pratt KP, Fujikawa K, Davie EW, Stoddard BL, Thompson AR: **Hemophilic factor VIII C1- and C2-domain missense mutations and their modeling to the 1,5-angstrom human C2-domain crystal structure.** *Blood* 2000, **96**:979-987.
101. Jacquemin M, Lavend'homme R, Benhida A, Vanzielegheem B, D'Oiron R, Lavergne J-M, Brackmann HH, Schwaab R, VandenDriessche T, Chuah MKL, Hoylaerts M, Gilles JGG, Peerlinck K, Vermeylen J, Saint-Remy J-MR: **A novel cause of mild/moderate haemophilia A: mutations scattered in the factor VIII C1 domain reduce factor VIII binding to von Willebrand factor.** *Blood* 2000, **96**:958-965.
102. Levinson B, Janco RL, Phillips JA III, Gitschier J: **A novel missense mutation in the factor VIII gene identified by analysis of amplified haemophilia DNA sequences.** *Nucleic Acids Res* 1987, **15**:9797-9805.
103. Jonsdottir S, Diamond C, Levinson B, Magnusson S, Jansson O, Gitschier J: **Missense mutations causing mild haemophilia A in Iceland detected by denaturing gradient gel electrophoresis.** *Hum Mutat* 1992, **1**(6):506-508.
104. Sukarova-Stefanovska E, Zisovski N, Muratovska O, Kostova S, Efremov GD: **Three novel point mutations causing haemophilia A.** *Haemophilia* 2002, **8**:715-718.
105. Koeberl DD, Bottema CD, Buerstedde JM, Sommer SS: **Functionally important regions of the factor IX gene have a low rate of polymorphism and a high rate of mutation in the dinucleotide CpG.** *Am J Hum Genet* 1989, **45**(3):448-457.
106. Onay UV, Kavakli K, Kiliç Y, Gürgey A, Aktuğlu G, Kemahli S, Ozbek U, Çağlayan SH: **Molecular pathology of haemophilia B in Turkish patients: identification of a large deletion and 33 independent point mutations.** *Br J Haematol* 2003, **120**(4):656-659.
107. Chu K, Wu SM, Stanley T, Stafford DW, High KA: **A mutation in the propeptide of Factor IX leads to warfarin sensitivity by a novel.** *J Clin Invest* 1996, **98**(7):1619-1625.
108. Heit JA, Thorland EC, Ketterling RP, Lind TJ, Daniels TM, Zapata RE, Ordonez SM, Kasper CK, Sommer SS: **Germline mutations in Peruvian patients with haemophilia B: pattern of mutation in AmerIndians is similar to the putative endogenous germline pattern.** *Hum Mutat* 1998, **11**(5):372-376.

109. Chen SH, Thompson AR, Zhang M, Scott CR: **Three point mutations in the factor IX genes of five haemophilia B patients. Identification strategy using localization by altered epitopes in their hemophilic proteins.** *J Clin Invest* 1989, **84**(1):113-118.
110. Wang NS, Zhang M, Thompson AR, Chen SH: **Factor IX Chongqing: a new mutation in the calcium-binding domain of factor IX resulting in severe haemophilia B.** *Thromb Haemost* 1990, **63**(1):24-26.
111. Espinós C, Casaña P, Haya S, Cid AR, Aznar JA: **Molecular analyses in haemophilia B families: identification of six new mutations in the factor IX gene.** *Haematologica* 2003, **88**(2):235-236.
112. Davis LM, McGraw RA, Ware JL, Roberts HR, Stafford DW: **Factor IX Alabama: a point mutation in a clotting protein results in haemophilia B.** *Blood* 1987, **69**(1):140-143.
113. Caglayan SH, Gökmen Y, Aktuglu G, Gurgey A, Sommer SS: **Mutations associated with haemophilia B in Turkish patients.** *Hum Mutat* 1997, **10**(1):76-79.
114. David D, Moreira I, Morais S, de Deus G: **Five novel factor IX mutations in unrelated haemophilia B patients.** *Hum Mutat* 1998, **1**:S301-S303.
115. Vidal F, Farssac E, Altisent C, Puig L, Gallardo D: **Factor IX gene sequencing by a simple and sensitive 15-hour procedure for haemophilia B diagnosis: identification of two novel mutations.** *Br J Haematol* 2000, **111**(2):549-551.
116. Noyes CM, Griffith MJ, Roberts HR, Lundblad RL: **Identification of the molecular defect in factor IX Chapel Hill: substitution of histidine for arginine at position 145.** *Proc Natl Acad Sci* 1983, **80**(14):4200-4202.
117. Aguilar-Martinez P, Romey MC, Schved JF, Gris JC, Demaille J, Claustres M: **Factor IX gene mutations causing haemophilia B: comparison of SSC screening versus systematic DNA sequencing and diagnostic applications.** *Hum Genet* 1994, **94**(3):287-290.
118. Liddell MB, Peake IR, Taylor SA, Lillcrap DP, Giddings JC, Bloom AL: **Factor IX Cardiff: a variant factor IX protein that shows abnormal activation is caused by an arginine to cysteine substitution at position 145.** *Br J Haematol* 1989, **72**(4):556-560.
119. Cargill M, Altshuler D, Ireland J, Sklar P, Ardlie K, Patil N, Shaw N, Lane CR, Lim EP, Kalyanaraman N, Nemesh J, Ziaugra L, Friedland L, Rolfe A, Warrington J, Lipshutz R, Daley GQ, Lander ES: **Characterization of single-nucleotide polymorphisms in coding regions of human genes.** *Nat Genet* 1999, **22**(3):231-238.
120. Bertina RM, van der Linden IK, Mannucci PM, Reinalda-Poot HH, Cupers R, Poort SR, Reitsma PH: **Mutations in haemophilia Bm occur at the Arg180-Val activation site or in the catalytic domain of factor IX.** *J Biol Chem* 1990, **265**(19):10876-10883.
121. Sakai T, Yoshioka A, Yamamoto K, Niinomi K, Fujimura Y, Fukui H, Miyata T, Iwanaga S: **Blood clotting factor IX Kashihara: amino acid substitution of valine-182 by phenylalanine.** *J Biochem* 1989, **105**(5):756-759.
122. Taylor SA, Liddell MB, Peake IR, Bloom AL, Lillcrap DP: **A mutation adjacent to the beta cleavage site of factor IX (valine 182 to leucine) results in mild haemophilia Bm.** *Br J Haematol* 1990, **75**(2):217-221.
123. David D, Rosa HA, Pemberton S, Diniz MJ, Campos M, Lavinha J: **Single-strand conformation polymorphism (SSCP) analysis of the molecular pathology of haemophilia B.** *Hum Mutat* 1993, **2**(5):355-361.
124. Ludwig M, Sabharwal AK, Brackmann HH, Olek K, Smith KJ, Birktoft JJ, Bajaj SP: **Haemophilia B caused by five different nondeletion mutations in the protease domain of factor IX.** *Blood* 1992, **79**(5):1225-1232.
125. Miyata T, Sakai T, Sugimoto M, Naka H, Yamamoto K, Yoshioka A, Fukui H, Mitsui K, Kamiya K, Umeyama H, Iwanaga S: **Factor IX Amagasaki: a new mutation in the catalytic domain resulting in the loss of both coagulant and esterase activities.** *Biochemistry* 1991, **30**(47):11286-11291.
126. Simioni P, Tormene D, Tognin G, Gavasso S, Bulato C, Iacobelli NP, Finn JD, Spiezia L, Radu C, Arruda VR: **X-linked thrombophilia with a mutant factor IX (factor IX Padua).** *N Engl J Med* 2009, **361**(17):1671-1675.
127. Chan V, Chan WW, Yip B, Chim CS, Chan TK: **Haemophilia B in a female carrier due to skewed inactivation of the normal X-chromosome.** *Am J Hematol* 1998, **58**(1):72-76.
128. Sugimoto M, Miyata T, Kawabata S, Yoshioka A, Fukui H, Takahashi H, Iwanaga S: **Blood clotting factor IX Niigata: substitution of alanine-390 by valine in the catalytic domain.** *J Biochem* 1988, **104**(6):878-880.
129. Attree O, Vidaud D, Vidaud M, Amselem S, Lavergne JM, Goossens M: **Mutations in the catalytic domain of human coagulation factor IX: rapid characterization by direct genomic sequencing of DNA fragments displaying an altered melting behavior.** *Genomics* 1989, **4**(3):266-272.
130. Ware J, Davis L, Frazier D, Bajaj SP, Stafford DW: **Genetic defect responsible for the dysfunctional protein: factor IX Long Beach.** *Blood* 1988, **72**(2):820-822.
131. DeLano WL: **The PyMOL Molecular Graphics System, Version 0.99.** DeLano Scientific San Carlos CA; 2002.
132. Markoff A, Gerke V, Bogdanova N: **Combined homology modelling and evolutionary significance evaluation of missense mutations in blood clotting factor VIII to highlight aspects of structure and function.** *Haemophilia* 2009, **15**(4):932-941.
133. Schwaab R, Ludwig M, Kochhan L, Oldenburg J, McVey JH, Egli H, Brackmann HH, Olek K: **Detection and characterisation of two missense mutations at a cleavage site in the factor VIII light chain.** *Thromb Res* 1991, **61**(3):225-234.
134. Hamaguchi M, Matsushita T, Tanimoto M, Takahashi I, Yamamoto K, Sugiura I, Takamatsu J, Ogata K, Kamiya T, Saito H: **Three distinct point mutations in the factor IX gene of three Japanese CRM + haemophilia B patients (factor IX BM Nagoya 2, factor IX Nagoya 3 and 4).** *Thromb Haemost* 1991, **65**(5):514-520.
135. Higuchi M, Wong C, Kochhan L, Olek K, Aronis S, Kasper CK, Kazazian HH Jr, Antonarakis SE: **Characterization of mutations in the factor VIII gene by direct sequencing of amplified genomic DNA.** *Genomics* 1990, **6**(1):65-71.
136. Frazier D, Smith KJ, Cheung WF, Ware J, Lin SW, Thompson AR, Reischer H, Bajaj SP, Stafford DW: **Mapping of monoclonal antibodies to human factor IX.** *Blood* 1989, **74**(3):971-977.
137. Bajaj SP: **Region of factor IXa protease domain that interacts with factor VIIIa: analysis of select haemophilia B mutants.** *Thromb Haemost* 1999, **82**(2):218-225.
138. Saunders RE, Perkins SJ: **CoagMDB: a database analysis of missense mutations within four conserved domains in five vitamin K-dependent coagulation serine proteases using a text-mining tool.** *Hum Mutat* 2008, **29**(3):333-344.
139. Song X, Geng Z, Zhu J, Li C, Hu X, Bian N, Zhang X, Wang Z: **Structure-function roles of four cysteine residues in the human arsenic (+3 oxidation state) methyltransferase (hAS3MT) by site-directed mutagenesis.** *Chem Biol Interact* 2009, **179**(2-3):321-328.
140. Waksman G, Kominos D, Robertson SC, Pant N, Baltimore D, Birge RB, et al: **Crystal structure of the phosphotyrosine recognition domain SH2 of v-src complexed with tyrosine-phosphorylated peptides.** *Nature* 1992, **358**:646-653.
141. Savas S, Kim DY, Ahmad MF, Shariff M, Ozcelik H: **Identifying functional genetic variants in DNA repair pathway using protein conservation analysis.** *Cancer Epidemiol Biomarkers Prev* 2004, **13**:801-807.

doi:10.1186/1423-0127-19-30

Cite this article as: C: *In Silico* profiling of deleterious amino acid substitutions of potential pathological importance in haemophilia A and haemophilia B. *Journal of Biomedical Science* 2012 19:30.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
www.biomedcentral.com/submit

